DIGITALCOMMONS @WAYNESTATE

Journal of Modern Applied Statistical Methods

Volume 4 | Issue 2 Article 12

11-1-2005

Quasi-Maximum Likelihood Estimation For Latent Variable Models With Mixed Continuous And Polytomous Data

Jens C. Eickhoff

University of Wisconsin – Madison, eickhoff@biostat.wisc.edu

Journal of Modern Applied Statistical Methods
November, 2005, Vol. 4, No. 2, 473- 481

1538 – 9472/05/$95.00

Quasi-Maximum Likelihood Estimation For Latent Variable Models With Mixed Continuous And Polytomous Data

Jens C. Eickhoff
Department of Biostatistics & Medical Informatics
University of Wisconsin – Madison

Latent variable modeling is a multivariate technique commonly used in the social and behavioral sciences. The models used in such analysis relate all observed variables to latent common factors. In many situations, however, some outcome variables are in polytomous form while other outcomes are measured on a continuous scale. Maximum likelihood estimation for latent variable models with mixed polytomous and continuous outcomes is computationally intensive and may become difficult to implement in many applications. In this article, a computationally practical, yet efficient, Quasi-Maximum Likelihood approach for latent variable models with mixed continuous and polytomous variables is proposed. Asymptotic properties of the estimator are discussed. Simulation studies are conducted to examine the empirical behavior and to compare it with existing methods.

Key words: multivariate analysis, polytomous outcome variables, Quasi-ML estimation.

Introduction

The problem of analyzing concepts or variables which are not directly observable and can only be measured through related indicators arises frequently in practice. In these situations, latent variable modeling provides a useful statistical technique. Statistical methods for analyzing covariances and other relationships between latent and observed variables were historically originated in psychometrics in the form of factor analysis which has later been extended to the more general structural equation analysis (Bentler, 1995; Bollen, 1989; Jöreskog and Sörbom, 1996). Today, latent variable models are extensively used in the behavioral and social sciences.

Most latent variable models are based on the assumption that the observed variables are continuous with a multivariate normal distribution. However, in many studies where data are obtained based on questionnaires, some or all observed outcome variables are typically in polytomous form. For example, data are frequently collected based on questionnaires with Likert scales (e.g., ``disagree", ``neutral", ``agree") responses. Because of its importance in many applications, there has been much attention in latent variable modeling involving polytomous outcomes and it remains an active area of research.

Bock and Lieberman (1970) considered a maximum likelihood method for factor analysis models with dichotomous outcome variables and only one factor. However, direct maximum likelihood analysis for models involving higher dimensional latent variables becomes computationally impractical because it requires maximization over multiple intractable integrals. This led to the development of multi-stage weighted least square estimation based on limited first and second-order sampling using polychoric and polyserial correlations (Muthén, 1984; Lee & Poon, 1987). Multi-stage weighted least squares (WLS) estimation procedures for structural equation models with polytomous outcome variables have been implemented in popular psychometrical software packages including LISCOMP (Muthén, 1987), EQS (Bentler, 1995), LISREL/PRELIS (Jöreskog &

Jens C. Eickhoff is an Associate Scientist in the Department of Biostatistics & Medical Informatics. He obtained his Ph. D. from Iowa State University. Email him at E-mail: eickhoff@biostat.wisc.edu.

Sörbom, 1996), and Mplus (Muthén & Muthén, 1998). These procedures, however, can experience problems of numerical instability, bias, non-convergence, and non-positive definiteness of weight matrices in situations of small sample sizes but large number of outcome variables (Reboussin & Liang, 1998). Sammel & Ryan (1997) and Shi & Lee (2000) used a Monte Carlo EM algorithm to perform maximum likelihood estimation in latent variables models with mixed discrete and continuous outcome variables. These procedures are computationally intensive as each E-step is approximated by Monte Carlo integration and no closed-form expressions are available in the M-steps. Moreover, many iterations are typically required to achieve convergence.

In this article, a computationally practical, yet efficient, Quasi-ML estimation procedure is proposed for factor analysis and structural equation models with mixed continuous and polytomous outcome variables. Asymptotic properties and standard error estimation are discussed. The Quasi-ML estimation can be easily implemented and does not require intensive computations. Simulation studies indicate that the proposed Quasi-ML estimator is substantially more efficient than traditional multi-stage WLS estimators, especially for models where the number of continuous outcome variables exceeds the number of polytomous outcomes.

This article is organized as follows. In the Methodology section, the general model and motivation for the proposed approach, as well as the Quasi-ML estimation procedure and the computation of asymptotic standard errors are described. The results of a simulation study, where the performance of the proposed Quasi-ML estimation is compared with traditional multi-stage weighted least square estimation techniques, is presented in the Results section. Finally, a brief conclusion is given in the last section.

Methodology

Consider a multivariate mixed-type variable situation with p_1 continuous and p_2 polytomous outcome variables and n observations. Let $y_i = (y_{1i}, \cdots, y_{p_1 i})'$ denote the set of continuous outcome variables and $z_i = (z_{1i}, \cdots, z_{p_2 i})'$ denote the set of polytomous outcome variables, each with $c(k)$ categories $(k = 1, \cdots, p_2)$, measured on the i^{th} individual. To motivate the model, assume that the set of continuous and polytomous outcome variables can be explained by a smaller number of $q\ (q < p_1 + p_2)$ unobserved latent variables $f_i = (f_{1i}. \cdots, f_{qi})'$. For ease of notation, a measurement or confirmatory factor analysis model is considered as follows. The notation can be easily extended to utilize the more general structural equation model framework. The standard linear measurement model for the continuous outcome variables for the i^{th} observation can be expressed as

$$y_i = \mu + \mathbf{\Lambda}\, f_i + \varepsilon_i, \quad i = 1, \cdots, n, \qquad (1)$$

where ε_i is a vector of measurement errors and the parameters μ and $\mathbf{\Lambda}$ contain some restricted elements. It is assumed that

$$f_i \sim N(\mu_f, \mathbf{\Sigma}_f),$$
$$\varepsilon_i \sim N(0, \mathbf{\Psi}),$$

where the elements of $\mu_f, \mathbf{\Sigma}_f,$ and $\mathbf{\Psi}$ are unrestricted, free parameters. Furthermore, it is assumed that, conditional on f_i, the elements of y_i are independent, i.e., $\mathbf{\Psi}$ is set to be a diagonal matrix. Likewise, for the polytomous outcome variables, it is assumed that conditional on f_i, the elements of z_i are independent and that each z_{ki}, $(k = 1, \cdots, p_2)$ relates to the latent variables through a probit response probability function, i.e.,

$$P(z_{ki} \le c_j \mid f_i) = \mathbf{\Phi}(\alpha_{k_j} + \beta_k' f_i), \qquad (2)$$

for category c_j, $j = 1, \cdots, c(k) - 1$ and $\alpha_{k_1} < \cdots < \alpha_{k_{c(k)-1}}$. The intercept and slope parameters, α_{k_j} and β_k, describe the

measurement properties of the k^{th} polytomous outcome variable.

The model described by (1) and (2) contains the factor indeterminacy inherent in this type of latent variable models. That is, the same model can be expressed using transformed parameters and factors. To remove this indeterminacy, the following standard identification form (Wall & Amemiya, 2000) for sub-model (1) is used,

$$y_i = \begin{pmatrix} 0 \\ \mu_y \end{pmatrix} + \begin{pmatrix} I_q \\ \Lambda_y \end{pmatrix} f_i + \varepsilon_i, \quad i = 1, \cdots, n,$$

where μ_y is a $(p_1 - q)\times 1$ vector and Λ_y is a $(p_1 - q)\times q$ matrix with unrestricted parameters. If $q > p_1$, additional measurement parameters in sub-model (2) are restricted. Note that this is an interpretable and meaningful identification parameterization which allows for assessing latent variable characteristics because parameters corresponding to the latent variables, i.e., μ_f and Σ_f, remain unrestricted. This is particularly useful in multi-group analysis situations where the main interest lies in the comparison of latent variable characteristics between different sampling groups, e.g., sex, gender, etc.

Quasi-Maximum Likelihood Estimation

Let $\boldsymbol{Y} = (y_1, \cdots, y_n)$ and $\boldsymbol{Z} = (z_1, \cdots, z_n)$ denote the observed data matrices from a random sample of the underlying population. Furthermore, denote the model parameters as,

$$\alpha = (\alpha_{1_1}, \cdots, \alpha_{1_{c(1)-1}}, \cdots, \alpha_{p_{2_1}}, \cdots, \alpha_{p_{2c(p_2)-1}})',$$
$$\beta = (\beta_1', \cdots, \beta_{p_2}')',$$

and

$$\theta_y = \left(\mu_y', (\text{vec}\,\Lambda_y)', (\text{vec}\,\Psi)'\right)',$$
$$\theta_z = \left(\alpha', (\text{vec}\,\beta)'\right)',$$
$$\theta_f = \left(\mu_f', (\text{vec}\,\Sigma_f)'\right)'.$$

The log-likelihood function based on the observed data is given by

$$l(\theta_y, \theta_z, \theta_f \mid \text{Y, Z}) = \log p(\text{Y}; \theta_y, \theta_f) + \log p(\text{Z} \mid \text{Y}; \theta_z, \theta_f). \quad (3)$$

Because $\log p(\boldsymbol{Z} \mid \boldsymbol{Y}; \theta_z, \theta_f)$ involves multiple integration which cannot be evaluated in closed form, direct maximization of this log-likelihood function is impractical. Various approaches have been proposed to overcome this computational burden. Sammel & Ryan (1997) and Shi & Lee (2000) proposed utilizing a Monte Carlo EM estimation approach. However, the EM algorithm is known to be slow and may require many iterations to achieve convergence. Moreover, the M-step in these approaches requires iterative procedures which might be time consuming, especially in models involving many polytomous outcomes.

The Quasi-ML approach (Besag, 1975) has become a popular tool in situations where the true likelihood function is computationally intractable but can be approximated by a function that is easier is evaluate. Quasi-ML methods may not always yield efficient estimators but they are usually consistent as long as the first derivatives of the quasi likelihood function has mean 0 at the true parameter values (Le Cessie & Houwelingen, 1994). In the following, a Quasi-ML approach is proposed where the second term of the right hand side of the log-likelihood function in (3) is approximated by a function which is computationally easy to evaluate. Specifically, the Quasi- log-likelihood for the i^{th} observation is expressed as

$$l_i^p = \log p(y_i; \theta_y, \theta_f) + \sum_{k=1}^{p_2} p(z_{ki} \mid y_i; \theta_z, \theta_f),$$

where $p(y_i; \theta_y, \theta_f)$ is a multivariate normal density function with mean

$$\mu(\theta_y, \theta_f) = (\mu_f', (\mu_y + \Lambda_y \mu_f)')'$$

and covariance matrix

$$\boldsymbol{\Sigma}(\theta_y, \theta_f) = \begin{pmatrix} I_q \\ \boldsymbol{\Lambda}_y \end{pmatrix} \boldsymbol{\Sigma}_f \begin{pmatrix} I_q & \boldsymbol{\Lambda}_y \end{pmatrix} + \boldsymbol{\Psi}.$$

Standard evaluation of the conditional distribution, $z_{ki} \mid y_i$, leads to

$$P(z_{ki} \le c_j \mid y_i; \theta_y, \theta_f) = \boldsymbol{\Phi}\left(\frac{\alpha_{k_j} + \beta_k' \mu_{f_i|y_i}}{\sqrt{1 + \beta_k' \boldsymbol{\Sigma}_{f_i|y_i} \beta_k}} \right),$$

where $1 \le k \le c(k) - 1$ and

$$\mu_{f_i|y_i} = \mu_f + \boldsymbol{\Sigma}_f \begin{pmatrix} I_q & \boldsymbol{\Lambda}_y \end{pmatrix} \left(\begin{pmatrix} I_q \\ \boldsymbol{\Lambda}_y \end{pmatrix} \boldsymbol{\Sigma}_f \begin{pmatrix} I_q \\ \boldsymbol{\Lambda}_y \end{pmatrix}' \right)^{-1} \left(y_i - \begin{pmatrix} \mu_f \\ \mu_y - \boldsymbol{\Lambda} \mu_f \end{pmatrix} \right),$$

$$\boldsymbol{\Sigma}_{f_i|y_i} = \boldsymbol{\Sigma}_f - \boldsymbol{\Sigma}_f \begin{pmatrix} I_q & \boldsymbol{\Lambda}_y \end{pmatrix} \left(\begin{pmatrix} I_q \\ \boldsymbol{\Lambda}_y \end{pmatrix} \boldsymbol{\Sigma}_f \begin{pmatrix} I_q \\ \boldsymbol{\Lambda}_y \end{pmatrix}' \right)^{-1} \begin{pmatrix} I_q \\ \boldsymbol{\Lambda}_y \end{pmatrix} \boldsymbol{\Sigma}_f.$$

The total Quasi log-likelihood is then the sum of the l_i^p's, i.e.,

$$\begin{aligned} l^p &= \sum_{i=1}^{n} l_i^p = \sum_{i=1}^{n} \log p(y_i; \theta_y, \theta_f) + \sum_{i=1}^{n} \sum_{k=1}^{p} p(z_{ki} \mid y_i; \theta_z, \theta_f) \\ &\propto -\frac{n}{2} \left(\log \left| \Sigma(\theta_y, \theta_f) \right| + \frac{n-1}{n} tr(\mathrm{S}_y \Sigma^{-1}(\theta_y, \theta_f)) + \right. \\ &\quad \left. \left(\bar{y} - \mu(\theta_y, \theta_f) \right)' \Sigma^{-1}(\theta_y, \theta_f) \left(\bar{y} - \mu(\theta_y, \theta_f) \right) \right) \\ &\quad + \sum_{i=1}^{n} \sum_{k=1}^{p} p(z_{ki} \mid y_i; \theta_z, \theta_f), \end{aligned} \tag{4}$$

where $\bar{y}$ is the sample mean, and $\boldsymbol{S}_y$ is the empirical covariance matrix of $y_i = (y_{1i}, \cdots, y_{p_i i})'$. Note that for a model with several continuous outcomes but only one polytomous outcome variable, the Quasi-log-likelihood function (4) is identical with the log-likelihood function (3).

The Quasi-ML estimator $(\hat{\theta}_y, \hat{\theta}_z, \hat{\theta}_f)$ is obtained by solving

$$\begin{aligned} &S(\theta_y, \theta_z, \theta_f) \\ &= \sum_{i=1}^{n} s_i(\theta_y, \theta_z, \theta_f) = \sum_{i=1}^{n} \frac{\partial l_i^p(\theta_y, \theta_z, \theta_f)}{\partial(\theta_y, \theta_z, \theta_f)} = 0. \end{aligned} \tag{5}$$

Explicit solutions for solving (5) are not available and therefore an iterative procedure is required. Because the number of parameters in (4) is usually relatively large, a derivative free optimization procedure as the Nelder-Mead simplex algorithm may not be computationally efficient. On the other hand, using an efficient optimization procedure such as the Newton-Raphson algorithm requires evaluation the first partial derivatives and the Hessian matrix which might be, due to the complexity of the objective function in (4), a tedious task. A good compromise is using a quasi Newton-Raphson algorithm with numerical derivatives which is easy to implement and numerically stable.

Standard Errors

For the computation of confidence intervals for the Quasi-ML parameter estimates, standard error estimates are required. A sandwich estimator can be used to estimate standard errors of Quasi-ML parameter estimates. It follows from the delta theorem that, under mild regularity conditions (see, e.g., Stuart and Ord, 1991), the distribution of $\sqrt{n}(\hat{\theta}_y - \theta_y, \hat{\theta}_z - \theta_z, \hat{\theta}_f - \theta_f)'$ converges to a $N(0, \boldsymbol{\Delta})$ distribution with

$$\boldsymbol{\Delta} = n \boldsymbol{I}^{-1} \boldsymbol{D} \boldsymbol{I}^{-1},$$

where

$$\begin{aligned} \boldsymbol{D} &= \mathrm{cov}\left(S(\theta_y, \theta_z, \theta_f) \right), \\ \boldsymbol{I} &= E\left(S(\theta_y, \theta_z, \theta_f) \right). \end{aligned}$$

Estimates of $\boldsymbol{D}$ and $\boldsymbol{I}$ can be obtained by computing

$$\hat{\boldsymbol{D}} = \sum_{i=1}^{n} s_i(\hat{\theta}_y, \hat{\theta}_z, \hat{\theta}_f) \left(s_i(\hat{\theta}_y, \hat{\theta}_z, \hat{\theta}_f) \right)' \tag{6}$$

and

$$\hat{\boldsymbol{I}} = -\sum_{i=1}^{n} \frac{\partial\, s_i(\hat{\theta}_y, \hat{\theta}_z, \hat{\theta}_f)}{\partial\,(\theta_y, \theta_z, \theta_f)'}. \tag{7}$$

Expressions (6) and (7) can be obtained using the numerical first and second order derivatives in the last iteration step of the quasi Newton-Raphson algorithm used to solve (5).

Starting Values

As the quasi Newton-Raphson algorithm used to solve (5) is an iterative procedure, starting values for the model parameters are required. One way to obtain starting values is to treat the sub-models (1) and (2) separately. Specifically, starting values for the parameters corresponding to sub-model (1) can be computed using standard estimation procedures for fitting latent variable models with continuous outcomes (Bollen, 1989). These estimates can be used to estimate factor scores, i.e.

$$\tilde{f}_i = \left(\begin{pmatrix} I_q \\ \tilde{\boldsymbol{\Lambda}}_y \end{pmatrix}' \tilde{\boldsymbol{\Psi}}^{-1} \begin{pmatrix} I_q \\ \tilde{\boldsymbol{\Lambda}}_y \end{pmatrix}\right)^{-1} \begin{pmatrix} I_q \\ \tilde{\boldsymbol{\Lambda}}_y \end{pmatrix}' \tilde{\boldsymbol{\Psi}}^{-1} \left(y_i - \begin{pmatrix} 0 \\ \tilde{\mu}_y \end{pmatrix}\right),$$

where $\tilde{\boldsymbol{\Lambda}}_y, \tilde{\boldsymbol{\Psi}}$, and $\tilde{\mu}_y$ are parameter estimates obtained using standard estimation procedures for latent variables models with continuous outcomes. The latent variable f_i of sub-model (2) can then be replaced by the factor scores $\tilde{f}_i$ and standard probit regression can be performed to obtain starting values for θ_z.

Results

The purpose of this simulation study is to compare the performance of the proposed Quasi-ML estimation approach with the traditional multi-stage WLS estimation approach which is currently considered the gold standard of fitting mixed latent variable models with continuous and polytomous outcomes. In the following, a confirmatory factor analysis model models with three continuous outcome variables and various numbers of polytomous outcome variables are considered. It is assumed that each polytomous outcome variable has three categories. Sub-model (1) is given by

$$\begin{pmatrix} y_{1i} \\ y_{2i} \\ y_{3i} \end{pmatrix} = \begin{pmatrix} 0 \\ \mu_{y_1} \\ \mu_{y_2} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \lambda_1 & \lambda_2 \end{pmatrix} \begin{pmatrix} f_{1i} \\ f_{2i} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \end{pmatrix},$$

where

$$\begin{pmatrix} f_{1i} \\ f_{2i} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{f_1} \\ \mu_{f_2} \end{pmatrix}, \begin{pmatrix} \sigma^2_{f_1} & \sigma_{f_1,f_2} \\ \sigma_{f_1,f_2} & \sigma^2_{f_2} \end{pmatrix}\right)$$

$i = 1, \cdots n$, and ε_{ki}, $k = 1,2,3$, are iid with $N(0, \psi^2)$ distribution. The parameters $\mu_{y_2}, \mu_{y_2}, \lambda_1, \lambda_2, \sigma^2_{f_1}, \sigma_{f_1,f_2}, \sigma^2_{f_2}$, and ψ^2 are unrestricted parameters with the true values $\mu_{y_2} = \mu_{y_3} = 1$, $\lambda_1 = \lambda_2 = 0.8$, $\sigma^2_{f_1} = \sigma^2_{f_2} = 1$, $\sigma_{f_1,f_2} = 0.5$, and $\psi^2 = 0.36$.

Sub-model (2), which corresponds to the polytomous outcome variables, each with three categories, is given by,

$$P(z_{ki} = c_j \mid f_1, f_2) = \begin{cases} \Phi(\alpha_{k_1} + \beta_{k1} f_1 + \beta_{k2} f_2), & j = 1 \\ \Phi(\alpha_{k_2} + \beta_{k1} f_1 + \beta_{k2} f_2) - \Phi(\alpha_{k_1} + \beta_{k1} f_1 + \beta_{k2} f_2), & j = 2 \end{cases},$$

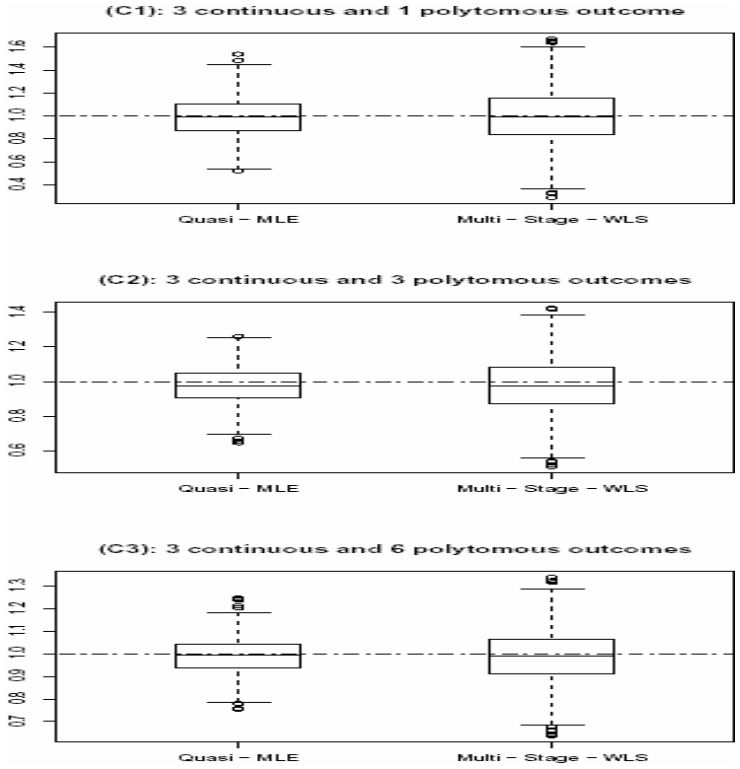
where $\alpha_{k_1}, \alpha_{k_2}, \beta_{k1}$, and β_{k2} are unrestricted parameters with true values $\alpha_{k_1} = 0.8$, $\alpha_{k_2} = 1.6$, $\beta_{k1} = 0.6$, and $\beta_{k2} = -0.6$. To facilitate generalization of the simulation results, the following three conditions on the number of polytomous outcome variables in the confirmatory factor models are considered:

(C1): Number of polytomous outcomes: 1

(C2): Number of polytomous outcomes: 3

(C3): Number of polytomous outcomes: 6

Figure 1: Boxpots for Quasi-ML and Multi-Stage WLS Estimators of $\sigma^2_{f_2}$ under Experimental Conditions (C1) – (C3) ($n = 500$)



Note that under experimental condition (C1), the Quasi-ML estimates are equivalent to the ML estimates. In order to compare the Quasi-ML estimation approach with the multi-stage WLS estimation approach, the model part corresponding to the polytomous outcome variables is first re-parameterized to the threshold model. This can be achieved by standardizing the intercept parameters $\alpha_{k_1}, \alpha_{k_2}$ to $\alpha^*_{k_1} = \alpha_{k_1} / \sqrt{1 - \beta' \boldsymbol{\Sigma}_f \beta} = 1$, $\alpha^*_{k_2} = \alpha_{k_2} / \sqrt{1 - \beta' \boldsymbol{\Sigma}_f \beta} = 2$, and the slope parameters β_{k1}, β_{k2} to $\beta^*_{k1} = \beta_{k1} / \sqrt{1 - \beta' \boldsymbol{\Sigma}_f \beta} = 0.75$ and $\beta^*_{k2} = \beta_{k2} / \sqrt{1 - \beta' \boldsymbol{\Sigma}_f \beta} = -0.75$, respectively.

The computation of the multi-stage WLS procedure was performed by using LISREL 8 and PRELIS 2. The Quasi-ML estimates were computed using R version 1.8.1.

The sample sizes considered were $n = 100$, $n = 500$, and $n = 1{,}000$. For each n and experimental condition (C1), (C2), and (C3), 1,000 simulations on samples were generated. The starting values for the Quasi-ML approach were computed as described in the previous section. Non-convergence was experienced in some cases for the multi-stage WLS approach when $n = 100$, especially for the model with 3 continuous and 6 polytomous outcomes (C3). For $n = 500$, the multi-stage WLS estimation procedure became numerically more stable. There were no convergence difficulties experienced for the Quasi-ML estimation for all sample sizes.

Figure 1 presents boxplots for the two estimators of the variance parameter $\sigma^2_{f_2}$ when $n = 500$, depicting the empirical distribution around the true parameter value $\sigma^2_{f_2} = 1.0$ under

Table 1: Empirical Bias and Root Mean Squared Error for Quasi-ML and Multi-Stage WLS Estimators for $\sigma^2_{f_2}$ under Experimental Conditions (C1) – (C3)

Experimental Condition	*n*		Quasi-MLE	Multi-Stage WLS
(C1)	100	Bias	0.044	0.054
		RMSE	0.142	0.220
	500	Bias	0.016	0.015
		RMSE	0.090	0.156
	1,000	Bias	0.010	0.008
		RMSE	0.052	0.120
(C2)	100	Bias	-0.010	-0.012
		RMSE	0.166	0.238
	500	Bias	0.026	0.023
		RMSE	0.110	0.165
	1,000	Bias	-0.009	0.011
		RMSE	0.079	0.118
(C3)	100	Bias	-0.081	0.022
		RMSE	0.199	0.244
	500	Bias	0.009	-0.007
		RMSE	0.131	0.155
	1,000	Bias	0.003	-0.001
		RMSE	0.102	0.129

experimental conditions (C1) – (C3). The general pattern given in Figure 1 can also been seen in boxplots for the other parameters and sample sizes. Table 1 gives the empirical bias and root mean squared error (RMSE) of the two estimators for the latent variable covariance parameters $\sigma^2_{f_1}$, σ_{f_2,f_2}, and $\sigma^2_{f_2}$. The cases where the multi-stage WLS estimator didn't converge were excluded when computing the empirical bias and RMSE.

The results indicate that the Quasi-ML estimator and the multi-stage WLS estimator are both unbiased for all coefficients and sample sizes. Under experimental conditions (C1) and (C2), the Quasi-ML estimate exhibit considerable less variability than the multi-stage WLS estimates. As the number of polytomous outcome variables increases this difference in RMSE between the two estimators becomes smaller. However, even under experimental condition (C3) (3 continuous and 6 polytomous outcomes), the Quasi-ML estimates still exhibit slightly less variability than the multi-stage WLS estimates.

Table 2 presents the empirical coverage probabilities of the nominal 95% confidence intervals for the Quasi-ML estimates of the latent variable covariance parameters $\sigma^2_{f_1}$, σ_{f_2,f_2}, and $\sigma^2_{f_2}$. The intervals were obtained by taking an estimate ± 1.96 times the corresponding estimated standard error. For all sample sizes, the constructed intervals give an empirical coverage close to the nominal level. Similar results were obtained for the other model parameters. Overall, the results indicate that the Quasi-ML standard errors can be used for valid statistical inference on the model parameters.

Conclusion

Multivariate polytomous data are common in psychosocial research. Consequently, there has been recently an increased interest in latent

Table 2: Empirical Coverage Probabilities for Quasi-ML estimates of Nominal 95% Confidence Intervals for Latent Variable Covariance Parameters

n	$\sigma^2_{f_1}$	σ_{f_2,f_2}	$\sigma^2_{f_2}$
100	91.2%	90.1%	90.9%
500	92.8%	91.3%	92.6%
1,000	94.0%	92.9%	93.9%

variable modeling involving polytomous outcome variables.

The parameter estimation of these types of models is computationally challenging. Traditional estimation techniques include multi-stage WLS procedures. However, it has been demonstrated that multi-stage WLS procedures can experience serious numerical problems, especially in situations of low prevalence, small sample sizes, or when fitting models with a large number of outcome variables.

Maximum likelihood estimation procedures have been proposed utilizing various types of EM algorithms (Sammel & Ryan, 1997; Shi & Lee, 2000). These procedures are numerically stable, yet computationally very intensive. In this article, a Quasi-ML method is proposed for parameter estimation of latent variable models with mixed continuous and polytomous variables. The procedure is computationally practical and can be easily implemented into standard statistical software (e.g., R, Splus, etc).

Simulation studies indicate that the proposed Quasi-ML estimator tends to be more efficient than traditional multi-stage WLS estimator, especially for models where the number of polytomous outcome variables is smaller than the number of continuous outcome variables. The Quasi-ML estimation of standard errors showed no substantial bias which warrants the performance of valid statistical inference. In summary, the proposed Quasi-ML estimation procedure appears to be efficient, computationally feasible, and a practical approach for latent variable models involving both continuous and polytomous outcomes.

References

Bentler, P. M. (1995). *EQS: Structural Equation Program Manual*. Los Angeles: BMDP Statistical Software.

Besag, J. (1975). The statistical analysis of non-lattice data. *Statistician, 24*, 179-195.

Bock, R. D. & Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika, 35*, 179-197.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Jöreskog, K. & Sörbom, D. (1996). *LISREL 8: User's Reference Guide*. Chicago: Scientific Software International.

Le Cessie, S. & *Van Houwelingen, J. C. (1994). Logistic regression for binary correlated data. Applied Statistics, 43*, 95-108.

Lee, S. Y. & Poon, W. Y. (1987). Two-step estimation of multivariate polychoric correlations. *Communications in Statistics: Theory and Methods, 16,* 307-320.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical and continuous latent variable indicators. *Psychometrika, 49,* 115-132.

Muthén, B. (1987) *LISCOMP, Analysis of linear structural equations with a comprehensive measurement model. Theoretical integration and uses's guide.* Mooresville, IN: Scientific Software.

Muthén, B. & Muthén, L. (1998). *Mplus User's Guide*. Los Angeles. CA: Muthen & Muthen.

Reboussin, B. A. & Liang, K. Y. (1998). An estimating equations approach for the LISCOMP model. *Psychometrika, 63*, 165-182.

Sammel, M. D. & Ryan, L. M. (1997). Latent variable models with mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society B, 59*, 667-678.

Shi, J. Q. & Lee, S. Y. (2000). Latent variable models with mixed continuous and polytomous data. *Journal of the Royal Statistical Society B, 62*, 77-87.

Stuart, A. & Ord, J.K. (1991). *Kendall's advanced theory of statistics*. London: Arnold.

Wall, M. M. & Amemiya, Y. (2000). Estimation for polynomial structural equation models. *Journal of the American Statistical Association, 95*, 929-940.